Article Watch, July 2012

Clive Slaughter

GHSU-UGA Medical Partnership, Athens, GA

This column highlights recently published articles that are of interest to the readership of this publication. We encourage ABRF members to forward information on articles they feel are important and useful to Clive Slaughter, GHSU-UGA Medical Partnership, 279 William St., Athens, GA 30607-1777. Phone: 706-369-5945; Fax: 706-369-5936; E-mail: cslaught@uga.edu; or to any member of the editorial board. Article summaries reflect the reviewer's opinions and not necessarily those of the association.

DNA SEQUENCING AND CHARACTERIZATION

Cherf G M, Lieberman K R, Rashid H, Lam C E, Karplus K, Akeson M. Automated forward and reverse ratcheting of DNA in a nanopore at 5-Å precision. *Nature Biotechnology* 30;2012:344–348.

Manrao E A, Derrington I M, Laszlo A H, Langford K W, Hopper M K, Gillgren N, Pavlenok M, Niederweis M, Gundlach J H. Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nature Biotechnology* 30;2012:349–353.

Two groups independently make progress toward realizing the 20-year-old idea of nanopore sequencing. The concept is to pass a DNA molecule through a protein nanopore and identify the bases by the ion current modulations that they cause as they process through this confining pore. The advantage of such a technique lies in the very long reads theoretically attainable. This contrasts with the short reads achieved by "next generation" methods. The promise, therefore, is the reduced need for computational reassembly, easier identification of indels, and improved analysis of repetitive sequences. The main impediment has hitherto been the high speed with which DNA is translocated through the pore by electrophoresis. Both of the present groups solve this problem by allowing a DNA polymerase to combine with the template, after hybridizing the template with an oligonucleotide, to place the speed of translocation under the control of the polymerase. The molecular pore is provided by α-hemolysin in the study of Cherf et al. and by the mycobacterial porin MspA in the study of Manrao et al. The next problem in need of solution is to identify the bases accurately. The present papers are of special interest, as their publication coincides with an announcement by Oxford Nanopore Technologies of the forthcoming launch of the first commercial nanopore sequencing device later this year (http://nanoporetech.com/news/press-releases/view/39). Although not yet published, the company claims to have sequenced the 48-kb genome of phage lambda as a single fragment using the technology.

doi: 10.7171/jbt.12-2302-007

Ding F, Manosas M, Spiering M M, Benkovic S J, Bensimon D, Allemand J-F, Croquette V. Single-molecule mechanical identification and sequencing. *Nature Methods* 9;2012:367–372.

"Next generation" sequencers rely on detecting fluorescently labeled nucleotides incorporated into amplified DNA templates by DNA polymerase or ligase. The limited read-length from which they suffer is caused, in part, by incomplete incorporation of nucleotide into the population of copies produced by amplification, which leads to out-of-phase signals that eventually overwhelm the true signal. "Third generation" sequencing, exemplified by the Pacific Biosystems sequencer, gets around this phasing problem by implementing true single-molecule sequencing. However, this incurs a higher error rate, owing partly to failure to detect fluorescence signals from single molecules as a result of rapid bleaching. Ding et al. here show proof-of-principle of an alternative, robust detection strategy based on measuring the length of a DNA hairpin tethered to a glass surface instead of detecting fluorescent signals. The free end of the DNA is conjugated to a magnetic bead. A magnetic field is then used to stretch the DNA so that it unfolds and becomes capable of hybridizing with a complementary oligonucleotide. After hybridization, the field is switched off. Any bound oligonucleotide blocks the refolding and changes the folded length of the hairpin. The beads are large enough to be imaged with a simple microscope and video camera, and their movement in and out of the plain of focus, caused by changes in the magnetic field, is accompanied by alterations in diffraction rings that allow the length of the DNA to be measured accurately. DNA templates can be identified by their

ability to hybridize to a specific probe, and they can be sequenced using a sequencing-by-ligation method similar to that used in the SOLiD sequencer. The read-out—progressive extension of the hairpin—is expected to provide greater accuracy than fluorescent signals for single-molecule sequencing.

CARBOHYDRATES AND GLYCOCONJUGATES

Liang C-W, Chang P-J, Lin Y-J, Lee Y-T, Ni C-K. High ion yields of carbohydrates from frozen solution by UV-MALDI. *Analytical Chemistry* 84;2012:3493–3499.

Much work has gone into improving the mass spectral performance of oligosaccharides in matrix-assisted laser desorption/ionization (MALDI). The present paper explores the use of matrix solutions frozen at 100 K in ultraviolet-assisted MALDI instead of using the conventional method of drying such solutions on the target. With the use of 2,5-dihydroxybenzoic acid as matrix dissolved in water-acetonitrile mixtures, samples of maltopentose and 12 K dextran are shown to be longer-lasting in the laser beam with the freezing method to generate ions that are subject to less postsource decay and to sustain ion generation efficiencies that are 20–30× greater than with conventional, dried droplet MALDI. How these improvements are brought about remains to be defined. A disadvantage of the method is the requirement for nonstandard cooling in the ion source.

Palmisano G, Melo-Braga M N, Engholm-Keller K, Parker B L, Larsen M R. Chemical deamidation: a common pitfall in large-scale N-linked glycoproteomic mass spectrometry-based analyses. *Journal of Proteome Research* 11; 2012:1949–1957.

In proteomic studies that attempt to screen for sites of N-linked glycosylation, protein N-glycosidase F (PNGase F) is commonly used to remove oligosaccharide. During this enzymatic reaction, glycan-linked asparagine residues are converted to aspartic acid. The replacement of asparagine by aspartic acid is often used as evidence of glycosylation. Palmisano et al. here draw attention to the need to control for nonenzymatic deamidation in studies that identify sites of glycosylation in this way. The authors use the absence of N-linked glycosylation in *Escherichia coli* to provide a negative control. They isolate membrane proteins from *E. coli* by carbonate membrane extraction, digest with PNGase F, and count the number of tryptic peptides in which an asparagine residue is converted to aspartic acid. The total number they discover is 391. Twenty-four of these sites become labeled with ^{18}O when $H_2{}^{18}O$ is present during PNGase F digestion, indicating that these particular sites undergo deamidation during experimental processing. Many sites identified in the study contain the anticipated Asn–Gly sequence, which undergoes deamidation through aspartamide rearrangement, a process that is alkali-catalyzed. This work highlights the need to perform suitable control experiments to avoid misidentification of N-liked glycosylation sites.

Yang S J, Zhang H. Glycan analysis by reversible reaction to hydrazide beads and mass spectrometry. *Analytical Chemistry* 84;2012:2232–2238.

Hydrazide-coated beads are used in this study to purify glycosylated peptides for subsequent analysis. The reducing ends of glycans released from glycoproteins conjugate with the beads to form hydrazones. Nonglycan moieties are washed away. The glycans are released for subsequent mass spectrometric analysis by treatment with acid. The procedure is applied here to study serum proteins. The method is expected to simplify glycan isolation and improve yields.

SMALL MOLECULE ANALYSIS AND METABOLOMICS

Causon T J, Cortes H J, Shellie R A, Hilder E F. Temperature pulsing for controlling chromatographic resolution in capillary liquid chromatography. *Analytical Chemistry* 84;2012:3362–3368.

Changes in temperature can have significant effects on the selectivity in liquid chromatography. The present work explores the use of rapid temperature changes to affect resolution of certain analytes, while leaving the chromatographic behavior of others unaltered. Such experiments would be impossible with columns of conventional diameter (2.1–4.6 mm i.d.), as radial temperature gradients would arise with temperature jumps, but with capillary columns (100 μm i.d.), rapid temperature jumps are practicable with resistive heating. They are used here in a manner analogous to programmed variation in mobile-phase composition. Such programming is demonstrated for ion-exchange chromatography of small anions in packed columns and for hydrophilic interaction chromatography of purines, pyrimidines, and nucleosides in silica monolith columns. Positive and negative temperature excursions are used. Resolution is demonstrated for analytes that coelute in isothermal separations. The methodology is proposed as a way of fine-tuning separations.

MASS SPECTROMETRY

Dupré M, Cantel S, Martinez J, Enjalbal C. Occurrence of C-terminal residue exclusion in peptide fragmentation by ESI and MALDI tandem mass spectrometry. *Journal of The American Society for Mass Spectrometry* 23;2012: 330–346.

Dupré et al. draw attention to an unexpectedly common peptide rearrangement that occurs during collision-induced dissociation. They study this rearrangement in a

set of synthetic peptides using low-energy collisions in a quadrupole time-of-flight (TOF) mass spectrometer and high-energy collisions in a TOF/TOF instrument. The rearrangement consists of the loss of the C-terminal residue (even a basic C-terminal residue), and it occurs in singly protonated (or cationized) peptides that have an internal arginine or histidine residue. The product may become the base peak in the fragment ion spectrum. It represents a $(b_{n-1}+H_2O)$ ion and is produced through an anhydride intermediate. The rearrangement is suppressed when the α-carboxyl group on the C-terminus is amidated. Importantly, when sequencing a peptide de novo, loss of the C-terminal residue can be confused with loss of the N-terminal residue, i.e., a y_{m-1} ion. Investigators doing de novo sequencing should be aware of this pitfall. The rearrangement is particularly relevant when proteolytic enzymes other than trypsin are used for protein digestion, as they are apt to generate peptides with internal arginine residues.

PROTEINS—PURIFICATION AND CHARACTERIZATION

Zakeri B, Fierer J O, Celik E, Chittock E C, Schwarz-Linek U, Moy V T, Howarth M. Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proceedings of the National Academy of Sciences USA* 109;2012:E690–E697.

Zakeri et al. have developed a method for stabilizing the association between transiently interacting molecules by forming a covalent bond without using chemical modification, artificial amino acids, or cysteines. Invasive strains of the Gram-positive bacterium, *Streptococcus pyrogenes,* contain a fibronectin-binding protein (FbaB) that enables the bacteria to be phagocytosed by endothelial cells. FbaB contains a collagen adhesion domain (CnaB2) that is stabilized by spontaneous formation of an intramolecular isopeptide bond between a lysine residue and an aspartic acid residue. The authors here split this domain into two parts and perform some re-engineering to provide a 13-residue peptide and a 138-residue protein that retain the ability to become cross-bridged by an isopeptide bond. The bond is formed with high yield in vitro and inside *Escherichia coli* cells. Formation of the bond is specific, occurs within minutes, and is insensitive to experimental conditions, including the presence of detergents. It requires no special conditions besides mixing the components together. The peptide is used as a tag named SpyTag; the protein is named SpyCatcher. When an ICAM-1 fusion protein containing SpyTag near its N-terminus is expressed on a mammalian cell surface, it is shown to bind to and react with SpyCatcher conjugated to Alexa Fluor 555 at the cell surface. The system is anticipated to provide a targetable lock between proteins within cells and to stabilize novel protein architectures.

PROTEOMICS

McAlister G C, Russell J D, Rumachik N G, Hebert A S, Syka J E P, Geer L Y, Westphall M S, Pagliarini D J, Coon J J. Analysis of the acidic proteome with negative electron-transfer dissociation mass spectrometry. *Analytical Chemistry* 84;2012:2875–2882.

The conventional use of acidic solutions and use of positive ion mass spectrometry (MS) in proteomics discriminate against acidic protein sequences, negatively charged post-translational modifications, such as phosphorylation and sialylation, and acid-labile species, such as histidine phosphate. Although electrospray ionization can be performed in negative ion mode, and basic mobile phases can be used instead of acidic ones, collision-induced dissociation of anions yields a paucity of sequence information. Nevertheless, sequence information may be acquired for negative ions by electron-detachment dissociation, electron-photodetachment dissociation, or negative electron-transfer dissociation. Negative electron-transfer dissociation is implemented here on a dual-cell quadrupole ion trap-orbitrap hybrid mass spectrometer and supplemented with high pH peptide separation. The methodology is used for study of the yeast proteome. The new workflow yields several thousands of peptide spectral matches, a larger number than any study has hitherto accomplished by any negative ion method. Previously unseen portions of the proteome are rendered observable. However, the results still compare unfavorably with positive ion MS, which produces some three times more matches. Further improvements in the methodology are anticipated to depend on improving negative electrospray performance, for example, by optimizing mobile-phase composition. Alternatively, use might be made of concurrent or supplemental ion activation for fragment formation.

FUNCTIONAL GENOMICS AND PROTEOMICS

MacArthur D G, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, Jostins L, Habegger L, Pickrell J K, Montgomery S B, Albers C A, Zhang Z D, Conrad D F, Lunter G, Zheng H, Ayub Q, DePristo M A, Banks E, Hu M, Handsaker R E, Rosenfeld J A, Fromer M, Jin M, Mu X J, Khurana E, Ye K, Kay M, Saunders G I, Suner M-M, Hunt T, Barnes I H A, Amid C, Carvalho-Silva D R, Bignell A H, Snow C, Yngvadottir B, Bumpstead S, Cooper D N, Xue Y, Romero I G, Genomes Project C, Wang J, Li Y, Gibbs R A, McCarroll S A, Dermitzakis E T, Pritchard J K, Barrett J C, Harrow J, Hurles M E, Gerstein M B, Tyler-Smith C. A systematic survey of loss-of-function variants in human protein-coding genes. *Science* 335;2012:823–828.

How many loss-of-function variants is the genome of an individual likely to contain? How many such variants are likely to be deleterious and how many less deleterious, because they affect functionally redundant genes? How many such variants are genuine, and how many are likely to represent sequencing artifacts? To answer these questions, MacArthur et al. survey 185 individuals whose genomes have been sequenced as part of the 1000 Genomes Project. They identify 2951 candidate variants in protein-coding regions that might cause loss of function, but then impose stringent filtering criteria, such as the presence of stop codons or indels affecting reading frame to ensure that loss of function is highly likely. They also exclude variants associated with compensatory mutations. Finally, they subject the list of candidate variants to stringent informatic and experimental validation to exclude sequencing artifacts. This filtering leaves 43.5% (1285) of the original candidates. Any one individual is estimated to bear ~100 loss-of-function variants judged to be genuine on these criteria. Approximately 20 are likely to be present in a homozygous state. Most are presumed to be common variants in nonessential genes. Nevertheless, the list includes 26 variants that are known disease-causing alleles, and 21 that are predicted to cause disease. This suggests that variants with large phenotypic effects are collectively numerous, although each individual one might be rare. This study presents a direct estimate of genetic load of deleterious genes in protein-coding regions. The implications are of concern to clinical geneticists seeking to estimate disease risk from individual genome-sequencing studies.

Shibata Y, Kumar P, Layer R, Willcox S, Gagan J R, Griffith J D, Dutta A. Extrachromosomal microDNAs and chromosomal microdeletions in normal tissues. *Science* 336; 2012:82–86.

The authors describe a new DNA entity in normal tissues and in cell lines: extrachromosomal, circular DNA species of 80–2000 bp in length, most commonly within the 200- to 400-bp range. Sequencing of these microDNAs indicates that they arise from protein-coding genes and map to unique chromosomal locations. They may arise during replication repair or homologous repair or else represent displaced Okazaki fragments from replication forks. The authors propose the theory that these microDNAs represent the DNA excised during the formation of chromosomal microdeletions. Such microdeletions are mosaically distributed in somatic cells and also arise in the germline. The authors also point out that microDNAs might contribute to functional differences between individual cells within a tissue and could mediate non-Mendelian genetic phenomena.

Patwardhan R P, Hiatt J B, Witten D M, Kim M J, Smith R P, May D, Lee C, Andrie J M, Lee S-I, Cooper G M, Ahituv N, Pennacchio L A, Shendure J. Massively parallel functional dissection of mammalian enhancers in vivo. *Nature Biotechnology* 30;2012:265–270.

Patwardhan et al. have devised a massively parallel in vivo reporter assay to measure the effects of sequence variation in *cis*-acting enhancer elements. They study three enhancers, for each one synthesizing a library composed of 100,000 mutants that individually differ from the WT sequence by 2–3% of residues. Each mutant is linked to a unique 20-nucleotide reporter tag in its transcriptional cassette. The entire ensemble of mutants is injected into a mouse tail-vein, and 24 h later, the transcriptional activity of the mutants in the liver is measured by massively parallel sequencing to determine the abundance of the transcribed reporter tags to which they are linked. The library covers all possible single-nucleotide substitutions of the targeted enhancer, and enhancers are found to be robust to changes in the sequence: the effects of most mutations are small, and only ~22% show greater than 1.2× change in activity. Most combinations of mutations are additive in effect. As expected, effect sizes are strongly dependent on the position of the mutation within the enhancer. The methodology is presently limited by the enhancers being episomal in nature so that they function outside of the normal controls of the nuclear environment. Nevertheless, the approach provides a convenient, high-throughput way of dissecting the effects of sequence variation in noncoding regulatory elements.

Bonn S, Zinzen R P, Girardot C, Gustafson E H, Perez-Gonzalez A, Delhomme N, Ghavi-Helm Y, Wilczynski B, Riddell A, Furlong E E M. Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nature Genetics* 44;2012:148–156.

How does chromatin modification affect promoter and enhancer use during the evolving program of cell-fate transitions in a developing embryo? How do chromatin modifications affect the timing of enhancer, promoter, and gene activity? Bonn et al. address these questions in *Drosophila* embryos using tissue-specific expression of transgenes that encode fluorescently tagged nuclear proteins. Embryos are subjected to formaldehyde cross-linking to freeze the chromatin at a defined stage of development. Nuclei are then sorted from each tissue of interest by FACS. The batches of tissue- and developmental stage-specific chromatin derived by this protocol are subjected to chromatin immunoprecipitation to interrogate RNA polymerase II (Pol II) occupancy, histone H3 density, and specific chromatin modifications at promoters, gene bodies, *cis*-regulatory elements, and repressed regions. The results define conditions associated with dynamic enhancer use and Pol II occupancy. The methodology is expected to

provide a generally applicable tool for investigating the operation of regulatory transcriptional networks in developmental biology.

CELL BIOLOGY AND TISSUE ENGINEERING

Mattis J, Tye K M, Ferenczi E A, Ramakrishnan C, O'Shea D J, Prakash R, Gunaydin L A, Hyun M, Fenno L E, Gradinaru V, Yizhar O, Deisseroth K. Principles for applying optogenetic tools derived from direct comparative analysis of microbial opsins. *Nature Methods* 9;2012:159–172.

Optogenetics is a methodology for modulating the activity of neurons at selected locations and at selected times in complex central nervous systems by using light stimulation delivered from a point source. The light signals are transduced by opsins, which are light-responsive membrane channel proteins of bacterial origin. In optogenetic experiments, they are used to control action potentials by producing depolarizing currents, hyperpolarizing currents, or specified signal-transduction events. Different opsins respond to light of different wavelengths. The present article provides guidance on how to choose the best opsin for any particular neurobiological study. It provides data on opsin properties and how these properties elicit changes in neural activity. It compares depolarizing and hyperpolarizing opsins and documents the ability of depolarizing opsins to elicit spikes in pyramidal cells and hyperpolarizing opsins to inhibit action potentials. It is hoped that the comparisons between opsins and the description of experimental conditions to be used for each will facilitate the design, implementation, and interpretation of experiments involving optogenetic tools.

Strickland D, Lin Y, Wagner E, Hope C M, Zayner J, Antoniou C, Sosnick T R, Weiss E L, Glotzer M. Tulips: tunable, light-controlled interacting protein tags for cell biology. *Nature Methods* 9;2012:379–384.

The regulation of spatial proximity between protein constituents of signaling pathways (e.g., kinases and their substrates) is a fundamental process, by which signal pathway activation is regulated. Optogenetic tools offer an attractive way of exercising exogenous activation of signaling pathways in a localized manner and at a time of the investigator's choosing. The present paper describes a new light-inducible dimerization tag for such experiments that is based on the interaction between the light, oxygen, or voltage 2 (LOV2) domain of *Avena sativa* phototropin 1 and an engineered PDZ domain. Photoactivation causes a conformational change in LOV2 that exposes a peptide to binding by the PDZ protein. The masking of the peptide before activation, the intrinsic affinity of the PDZ molecule for the peptide, and the lifetime of the photoexcited state can all be tuned by mutation to suit the sensitivities and response times needed for any particular study. The system is used to localize proteins to specific regions of yeast and mammalian cells and to trigger specific cell signaling pathways. It is shown to regulate nucleotide-exchange factors and scaffold proteins and kinases by recruitment to the plasma membrane. The methodology is potentially applicable to the control of diverse cellular regulatory processes, including transcriptional regulation and enzyme-substrate proximity.

IMAGING

Schnell U, Dijk F, Sjollema K A, Giepmans B N G. Immunolabeling artifacts and the need for live-cell imaging. *Nature Methods* 9;2012:152–158.

Immunofluorescence labeling requires cell or tissue fixation to ensure protein immobilization and ultrastructure preservation as well as permeabilization to allow antibody access to the targeted antigen. However, strong fixation may interfere with epitope recognition, and permeabilization may result in protein washout. In this paper, Schnell et al. document examples of protein extraction and relocalization in standard immunofluorescence protocols and use electron microscopy to show the nature and extent of ultrastructural damage produced during fixation and permeabilization. They argue that live-cell imaging of recombinant proteins linked to GFP should be used to complement immunofluorescence results to ensure reliability of the conclusions drawn.